that well represent the surface topography of the segment. The topographic features of interest (the bumps and holes) are now located within this representation of the surface. Initially, all the two-dimensional bumps and holes are identified within every cross-section. These two-dimensional features are then used in combination to construct their three-dimensional counterparts.

Exquisite specificity is characteristic of molecular docking. The representation of surface topography, as described accurately, conserves such specificity (see Fig. 2). Equally important, the set of discovered three-dimensional features is small enough to permit matching by exhaustive trial between complementary types.

Received for publication 4 May 1985.

REFERENCES

1. Crippen, G. M. 1978. The tree structural organization of proteins. *J. Mol. Biol.* 126:315–332.
2. Rose, G. D. 1979. Hierarchic organization of domains in globular proteins. *J. Mol. Biol.* 134:447–470.
3. Yuschok, T. J., and G. D. Rose. 1983. Hierarchic organization of globular proteins, a control study. *Int. J. Pep. Protein Res.* 21:479–484.
4. Smith, W. W., R. M Burnett, G. D. Darling, and M. L. Ludwig. 1977. Structure of the semiquinone form of flavodoxin from Clostridium MP; extension of 1.8 Å resolution and some comparisons with the oxidized state. *J. Mol. Biol.* 117:195–225.
5. Ptitsyn, O. B., and A. A. Rashin. 1975. A model of myoglobin self-organization. *Biophys. Chem.* 3:1–20.
6. Warshel, A., and M. Levitt. 1976. Folding and stability of helical proteins: carp myogen. *J. Mol. Biol.* 106:421–437.
7. Richmond, T. J., and F. M Richards. 1978. Packing of α-helices: geometric constraints, and contact areas. *J. Mol. Biol.* 119:537–555.
8. Greer, J., and B. L. Bush. 1978. Macromolecular shape, and surface maps by solvent exclusion. *Proc. Natl. Acad. Sci. USA.* 75:303–307.
9. Wodak, S. J., and J. Janin. 1978. Computer analysis of protein–protein interaction. *J. Mol. Biol.* 124:323–342.
10. Lee, R. H., and G. D. Rose. 1985. Molecular recognition: I. Automatic identification of topographic surface features. *Biopolymers.* 24:1613–1627.
11. Richards, F. M. 1977. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.
12. Connolly, M. L. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science (Wash. DC).* 221:709–713.

SIMPLE CHARGE-MEDIUM INTERACTION MODELS OF AMPIPHILIC PROTEINS WITH UNKNOWN TERTIARY STRUCTURE

ALEKSANDER BLUM
International Minerals and Chemical Corporation, Terre Haute, Indiana 47808

Many secreted proteins have unresolved crystallographic structure. Secondary structure can be predicted by several methods, but tertiary organization is unknown. Tertiary organization, with the resultant solvent surface contours, is the most important determinant of function, stability, and bioactivity of proteins. Electrostatic effects dominate the protein surface and its interaction with the surrounding layer of water molecules (1). Double basic residues (R,K) serve as recognition signals for post-translational modification of preproteins and as anchor points for trypsin-like enzymatic cleavage (2).

In a particular protein the pattern of charged, flexible, and polarizable side chains superimposed on secondary structure of polypeptide backbone is unique. Such patterns can be represented by superposition of charge/dipol clusters on the hydropathic profiles (3) (amino acid sequences oriented along a protein's amphiphilic axis). Examples of the cluster patterns for three different globular proteins, shown in Figs. 1–3, reveal substantial differences in their electrostatic environment. Hydrophilic charged microregions are sharp and dense for growth hormone; spotty and diffuse for ribonuclease; and absent for ubiquitin. However the average charge density is similar for all three proteins.

Small globular proteins devoid of quaternary structure show the largest effect of electrostatic interactions on protein conformation. In the early stages of protein folding these interactions can be seen in backbone-side residue interactions (4), and in later stages in solvent-protein surface behavior (aggregation, salt effects)(5). We are attempting to use the charge/dipol protein pattern in computer modeling to predict gross, general features of tertiary structure.

The principal feature of the model consists of substitution of amino acids for atoms in a computer program designed for modeling small molecules with up to 2,000 atoms. The schematic protein model can then be visually manipulated on the screen. The amino acid sequence of globular protein can be represented by a chain of cubic elements (amino acid residues) separated by 4–6 Å (6). The amino acid residue elements are defined as "bulk elastic points" with a short list of parameters (hydrophobicity, fractional charge, pK, side chain flexibility and length, solvent static accesibility, volume, α or β propensity) averaged by the moving window technique. The model is built within known sterical constraints for α-helix, β-sheet, and β-strand, taking advantage of flexibility of

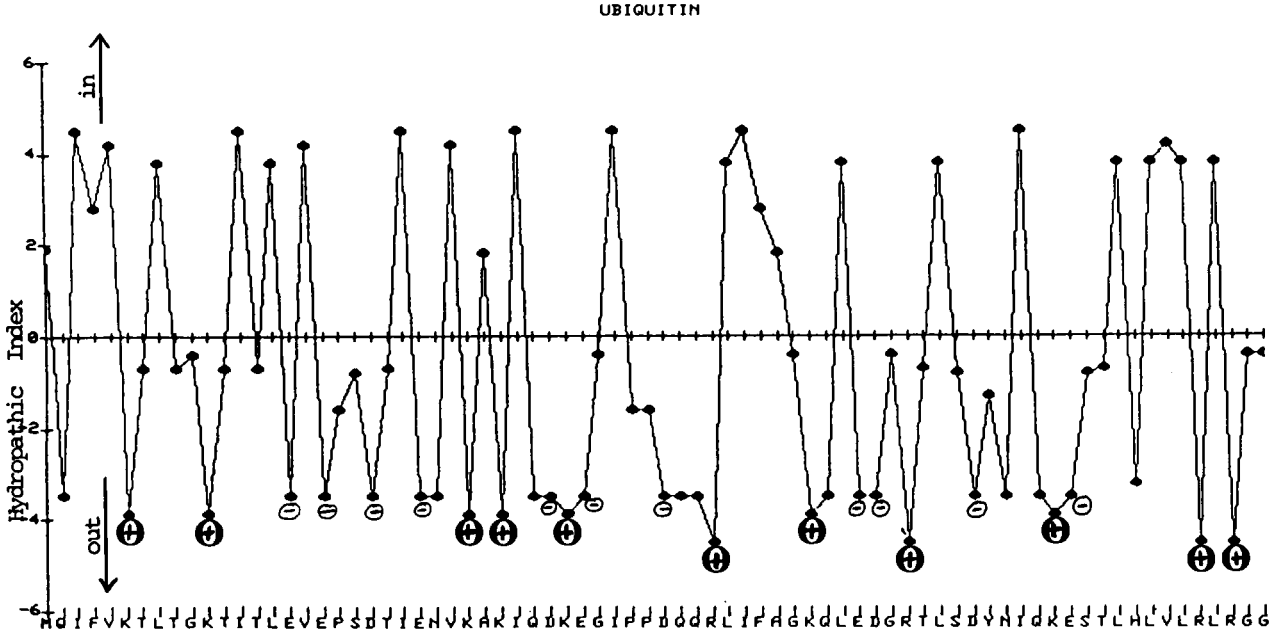

FIGURE 1 Charge pattern for ubiquitin.

long side chains for the D, E, K, R, G, and N residues. Visual manipulation of exposed charged regions based on minimalization of interaction energy within the sterical constraints of proteins helps to predict interactions between different regions of the protein.

This procedure can be used for two different proteins, protein-solvent, or protein cofactor complexes. Model predictions can be tested by comparison with the protein structure, if known, (e.g., ribonuclease), or by selective chemical modifications removing specific charges (deamidation, carboxymethylation, etc.). In vivo, the enzymatic charge-modification activity is concentrated in secretory tissue and facilitates the packaging of secretory proteins inside granules (7).

This model is useful for visualization of three-demensional best fit problems related to conformation; aggregation; protein-solvent or protein-receptor interaction; protein-antibody complex formation; interpretation of fluorescence spectra; chemical modification; active site prediction; and structure comparison within given protein subclasses.

While this model cannot be expected to predict precise details of tertiary structure at atomic resolution, it is simple, flexible, and may prove useful for gross prediction of protein tertiary structure. This type of model can also serve as a guide for appropriate chemical modifications, site-specific mutagenesis, and other experiments requiring knowledge of tertiary structure.

Received for publication 22 May 1985.

REFERENCES

1. Perutz, M. F. 1978. Electrostatic effects in proteins. *Science (Wash. DC)*. 201:1187–1191.
2. Docherty, K., and D. F. Steiner. 1982. Postranslational proteolysis in polypeptide hormone biosynthesis. *Annu. Rev. Physiol.* 44:625–638.
3. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.
4. Jerrigan, R. L., S. Miyazawa, and S. C. Szu. 1980. Stabilization of regular conformational regions in proteins by intraregion electrostatic interactions. *Macromolecules*. 13:518–525.
5. Finney, J. L., B. J. Gellathy, I. C. Colton, and J. Goodfellow. 1980. Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. *Biophys. J.* 10:17–30.
6. Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.
7. Diliberto Jr., E. J., and J. Axelrod. 1974. Characterization and substrate specificity of a protein carboxymethylase in the pituitary gland. *Proc. Natl. Acad. Sci. USA.* 71:1701–1704.

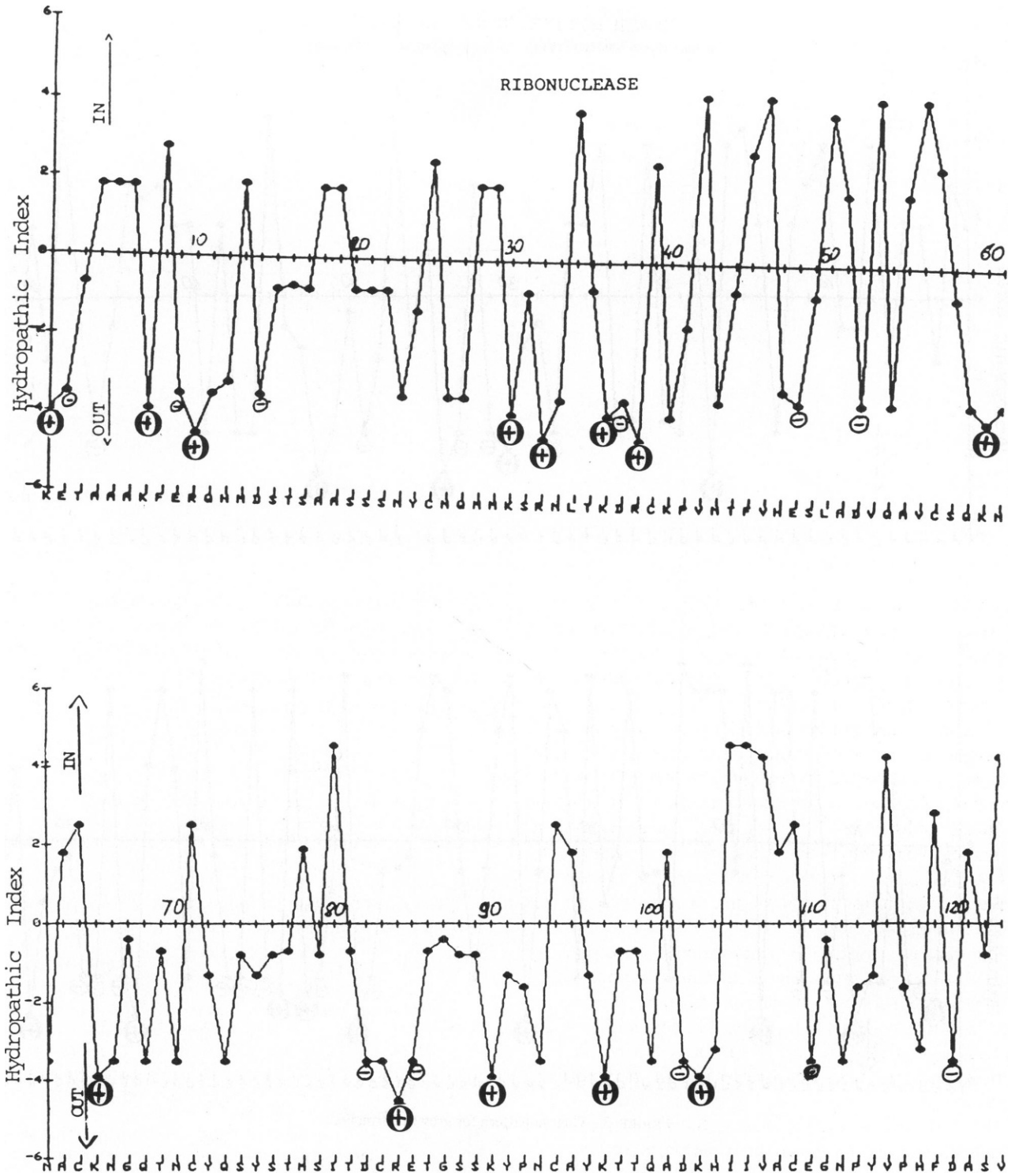

FIGURE 2 Charge pattern for ribonuclease.

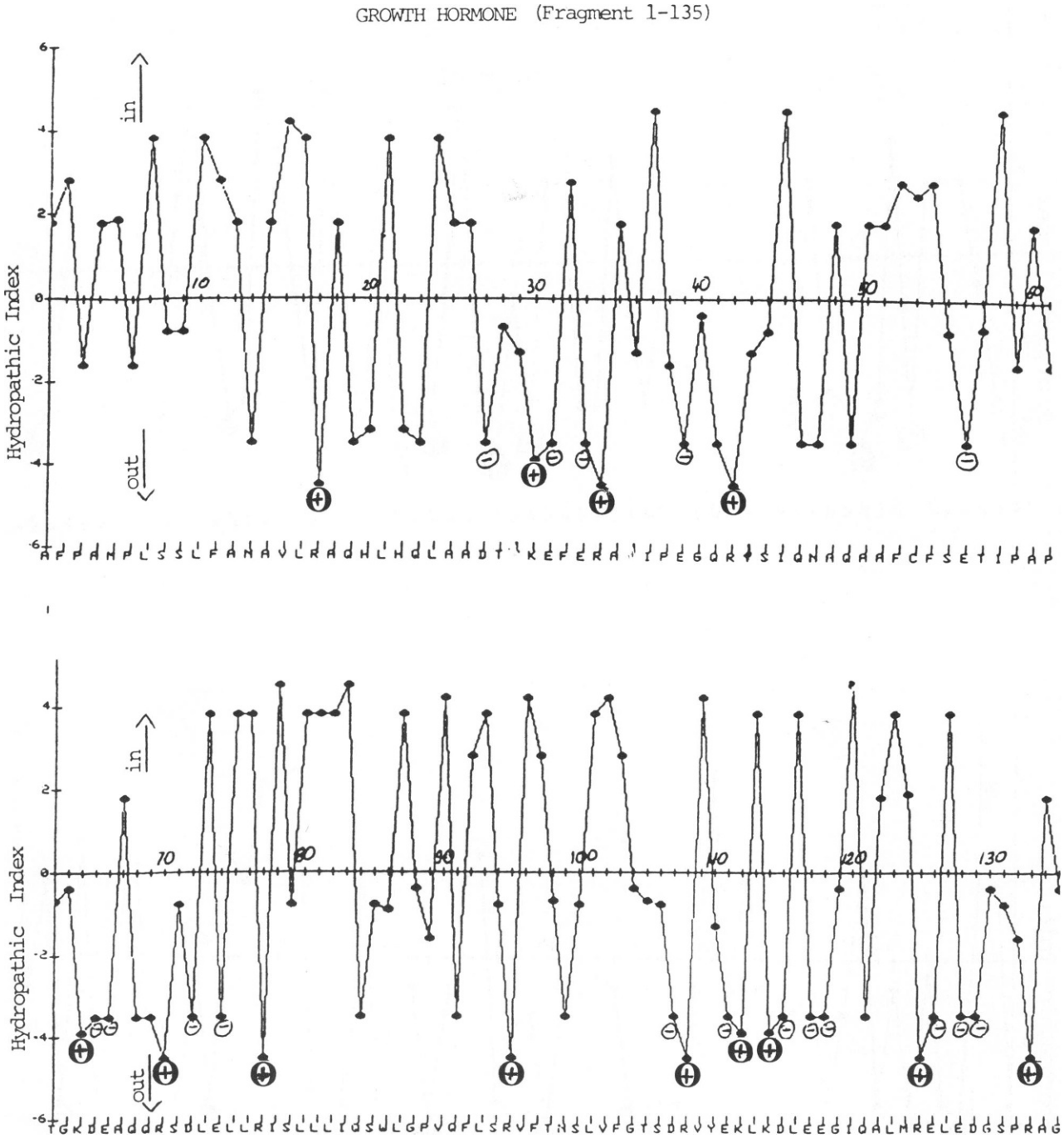

FIGURE 3 Charge pattern for growth hormone.